1812

PAGE: 1

RAW SEQUENCE LISTING
PATENT APPLICATION *US/08/826,361A*

DATE: 11/04/97
TIME: 14:59:21

INPUT SET: S21390.raw

> This Raw Listing contains the General
> Information Section and up to the first 5 pages.

```
1                               SEQUENCE LISTING
2
3   (1)     General Information:
4
5         (i) APPLICANT: Mosselman, Sieste
6                        Dijkema, Rein
7
8        (ii) TITLE OF INVENTION: Novel estrogen receptor
9
10      (iii) NUMBER OF SEQUENCES: 28
11
12       (iv) CORRESPONDENCE ADDRESS:
13            (A) ADDRESSEE: Akzo Nobel Patent Dept.
14            (B) STREET: 1300 Piccard Drive, Suite 206
15            (C) CITY: Rockville
16            (D) STATE: Maryland
17            (E) COUNTRY: US
18            (F) ZIP: 20850
19
20        (v) COMPUTER READABLE FORM:
21            (A) MEDIUM TYPE: Floppy disk
22            (B) COMPUTER: IBM PC compatible
23            (C) OPERATING SYSTEM: PC-DOS/MS-DOS
24            (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
25
26       (vi) CURRENT APPLICATION DATA:
27            (A) APPLICATION NUMBER: US 08/826,361
28            (B) FILING DATE: 26-MAR-1997
29            (C) CLASSIFICATION:
30
31     (viii) ATTORNEY/AGENT INFORMATION:
32            (A) NAME: Gormley, Mary E.
33            (B) REGISTRATION NUMBER: 34,409
34
35       (ix) TELECOMMUNICATION INFORMATION:
36            (A) TELEPHONE: 301-948-7400
37            (B) TELEFAX: 301-948-9751
38
39
40
41   (2) INFORMATION FOR SEQ ID NO: 1:
42
43        (i) SEQUENCE CHARACTERISTICS:
44            (A) LENGTH: 1434 base pairs
45            (B) TYPE: nucleic acid
46            (C) STRANDEDNESS: double
```

RAW SEQUENCE LISTING
PATENT APPLICATION *US/08/826,361A*

INPUT SET: S21390.raw

```
47              (D) TOPOLOGY: linear
48
49        (ii) MOLECULE TYPE: cDNA
50
51
52
53        (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
54
55   ATGAATTACA GCATTCCCAG CAATGTCACT AACTTGGAAG GTGGGCCTGG TCGGCAGACC        60
56
57   ACAAGCCCAA ATGTGTTGTG GCCAACACCT GGGCACCTTT CTCCTTTAGT GGTCCATCGC       120
58
59   CAGTTATCAC ATCTGTATGC GGAACCTCAA AAGAGTCCCT GGTGTGAAGC AAGATCGCTA       180
60
61   GAACACACCT TACCTGTAAA CAGAGAGACA CTGAAAAGGA AGGTTAGTGG GAACCGTTGC       240
62
63   GCCAGCCCTG TTACTGGTCC AGGTTCAAAG AGGGATGCTC ACTTCTGCGC TGTCTGCAGC       300
64
65   GATTACGCAT CGGGATATCA CTATGGAGTC TGGTCGTGTG AAGGATGTAA GGCCTTTTTT       360
66
67   AAAAGAAGCA TTCAAGGACA TAATGATTAT ATTTGTCCAG CTACAAATCA GTGTACAATC       420
68
69   GATAAAAACC GGCGCAAGAG CTGCCAGGCC TGCCGACTTC GGAAGTGTTA CGAAGTGGGA       480
70
71   ATGGTGAAGT GTGGCTCCCG GAGAGAGAGA TGTGGGTACC GCCTTGTGCG GAGACAGAGA       540
72
73   AGTGCCGACG AGCAGCTGCA CTGTGCCGGC AAGGCCAAGA GAAGTGGCGG CCACGCGCCC       600
74
75   CGAGTGCGGG AGCTGCTGCT GGACGCCCTG AGCCCCGAGC AGCTAGTGCT CACCCTCCTG       660
76
77   GAGGCTGAGC CGCCCCATGT GCTGATCAGC CGCCCCAGTG CGCCCTTCAC CGAGGCCTCC       720
78
79   ATGATGATGT CCCTGACCAA GTTGGCCGAC AAGGAGTTGG TACACATGAT CAGCTGGGCC       780
80
81   AAGAAGATTC CCGGCTTTGT GGAGCTCAGC CTGTTCGACC AAGTGCGGCT CTTGGAGAGC       840
82
83   TGTTGGATGG AGGTGTTAAT GATGGGGCTG ATGTGGCGCT CAATTGACCA CCCCGGCAAG       900
84
85   CTCATCTTTG CTCCAGATCT TGTTCTGGAC AGGGATGAGG GGAAATGCGT AGAAGGAATT       960
86
87   CTGGAAATCT TTGACATGCT CCTGGCAACT ACTTCAAGGT TTCGAGAGTT AAAACTCCAA      1020
88
89   CACAAAGAAT ATCTCTGTGT CAAGGCCATG ATCCTGCTCA ATTCCAGTAT GTACCCTCTG      1080
90
91   GTCACAGCGA CCCAGGATGC TGACAGCAGC CGGAAGCTGG CTCACTTGCT GAACGCCGTG      1140
92
93   ACCGATGCTT TGGTTTGGGT GATTGCCAAG AGGGGCATCT CCTCCCAGCA GCAATCCATG      1200
94
95   CGCCTGGCTA ACCTCCTGAT GCTCCTGTCC CACGTCAGGC ATGCGAGTAA CAAGGGCATG      1260
96
97   GAACATCTGC TCAACATGAA GTGCAAAAAT GTGGTCCCAG TGTATGACCT GCTGCTGGAG      1320
98
99   ATGCTGAATG CCCACGTGCT TCGCGGGTGC AAGTCCTCCA TCACGGGGTC CGAGTGCAGC      1380
```

INPUT SET: S21390.raw

```
100
101   CCGGCAGAGG ACAGTAAAAG CAAAGAGGGC TCCCAGAACC CACAGTCTCA GTGA        1434
102
103   (2) INFORMATION FOR SEQ ID NO: 2:
104
105       (i) SEQUENCE CHARACTERISTICS:
106             (A) LENGTH: 1251 base pairs
107             (B) TYPE: nucleic acid
108             (C) STRANDEDNESS: double
109             (D) TOPOLOGY: linear
110
111       (ii) MOLECULE TYPE: cDNA
112
113
114
115       (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
116
117   ATGAATTACA GCATTCCCAG CAATGTCACT AACTTGGAAG GTGGGCCTGG TCGGCAGACC          60
118
119   ACAAGCCCAA ATGTGTTGTG GCCAACACCT GGGCACCTTT CTCCTTTAGT GGTCCATCGC         120
120
121   CAGTTATCAC ATCTGTATGC GGAACCTCAA AAGAGTCCCT GGTGTGAAGC AAGATCGCTA         180
122
123   GAACACACCT TACCTGTAAA CAGAGAGACA CTGAAAAGGA AGGTTAGTGG GAACCGTTGC         240
124
125   GCCAGCCCTG TTACTGGTCC AGGTTCAAAG AGGGATGCTC ACTTCTGCGC TGTCTGCAGC         300
126
127   GATTACGCAT CGGGATATCA CTATGGAGTC TGGTCGTGTG AAGGATGTAA GGCCTTTTTT         360
128
129   AAAAGAAGCA TTCAAGGACA TAATGATTAT ATTTGTCCAG CTACAAATCA GTGTACAATC         420
130
131   GATAAAAACC GGCGCAAGAG CTGCCAGGCC TGCCGACTTC GGAAGTGTTA CGAAGTGGGA         480
132
133   ATGGTGAAGT GTGGCTCCCG GAGAGAGAGA TGTGGGTACC GCCTTGTGCG GAGACAGAGA         540
134
135   AGTGCCGACG AGCAGCTGCA CTGTGCCGGC AAGGCCAAGA GAAGTGGCGG CCACGCGCCC         600
136
137   CGAGTGCGGG AGCTGCTGCT GGACGCCCTG AGCCCCGAGC AGCTAGTGCT CACCCTCCTG         660
138
139   GAGGCTGAGC CGCCCCATGT GCTGATCAGC CGCCCCAGTG CGCCCTTCAC CGAGGCCTCC         720
140
141   ATGATGATGT CCCTGACCAA GTTGGCCGAC AAGGAGTTGG TACACATGAT CAGCTGGGCC         780
142
143   AAGAAGATTC CCGGCTTTGT GGAGCTCAGC CTGTTCGACC AAGTGCGGCT CTTGGAGAGC         840
144
145   TGTTGGATGG AGGTGTTAAT GATGGGGCTG ATGTGGCGCT CAATTGACCA CCCCGGCAAG         900
146
147   CTCATCTTTG CTCCAGATCT TGTTCTGGAC AGGGATGAGG GGAAATGCGT AGAAGGAATT         960
148
149   CTGGAAATCT TTGACATGCT CCTGGCAACT ACTTCAAGGT TTCGAGAGTT AAAACTCCAA        1020
150
151   CACAAAGAAT ATCTCTGTGT CAAGGCCATG ATCCTGCTCA ATTCCAGTAT GTACCCTCTG        1080
152
```

RAW SEQUENCE LISTING
PATENT APPLICATION   *US/08/826,361A*

INPUT SET: S21390.raw

```
153    GTCACAGCGA CCCAGGATGC TGACAGCAGC CGGAAGCTGG CTCACTTGCT GAACGCCGTG    1140
154
155    ACCGATGCTT TGGTTTGGGT GATTGCCAAG AGCGGCATCT CCTCCCAGCA GCAATCCATG    1200
156
157    CGCCTGGCTA ACCTCCTGAT GCTCCTGTCC CACGTCAGGC ATGCGAGGTG A             1251
158
159
160
161    (2) INFORMATION FOR SEQ ID NO: 3:
162
163         (i) SEQUENCE CHARACTERISTICS:
164             (A) LENGTH: 66 amino acids
165             (B) TYPE: amino acid
166             (C) STRANDEDNESS: single
167             (D) TOPOLOGY: linear
168
169        (ii) MOLECULE TYPE: peptide
170
171
172
173        (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
174
175        Cys Ala Val Cys Ser Asp Tyr Ala Ser Gly Tyr His Tyr Gly Val Trp
176        1               5                   10                  15
177
178        Ser Cys Glu Gly Cys Lys Ala Phe Phe Lys Arg Ser Ile Gln Gly His
179                    20              25                  30
180
181        Asn Asp Tyr Ile Cys Pro Ala Thr Asn Gln Cys Thr Ile Asp Lys Asn
182                35                  40                  45
183
184        Arg Arg Lys Ser Cys Gln Ala Cys Arg Leu Arg Lys Cys Tyr Glu Val
185            50                  55                  60
186
187        Gly Met
188        65
189
190    (2) INFORMATION FOR SEQ ID NO: 4:
191
192         (i) SEQUENCE CHARACTERISTICS:
193             (A) LENGTH: 233 amino acids
194             (B) TYPE: amino acid
195             (C) STRANDEDNESS: single
196             (D) TOPOLOGY: linear
197
198        (ii) MOLECULE TYPE: peptide
199
200
201
202        (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
203
204        Leu Val Leu Thr Leu Leu Glu Ala Glu Pro Pro His Val Leu Ile Ser
205        1               5                   10                  15
```

INPUT SET: S21390.raw

```
206
207        Arg Pro Ser Ala Pro Phe Thr Glu Ala Ser Met Met Met Ser Leu Thr
208                     20              25              30
209
210        Lys Leu Ala Asp Lys Glu Leu Val His Met Ile Ser Trp Ala Lys Lys
211                     35              40              45
212
213        Ile Pro Gly Phe Val Glu Leu Ser Leu Phe Asp Gln Val Arg Leu Leu
214             50              55              60
215
216        Glu Ser Cys Trp Met Glu Val Leu Met Met Gly Leu Met Trp Arg Ser
217        65              70              75              80
218
219        Ile Asp His Pro Gly Lys Leu Ile Phe Ala Pro Asp Leu Val Leu Asp
220                     85              90              95
221
222        Arg Asp Glu Gly Lys Cys Val Glu Gly Ile Leu Glu Ile Phe Asp Met
223                     100             105             110
224
225        Leu Leu Ala Thr Thr Ser Arg Phe Arg Glu Leu Lys Leu Gln His Lys
226             115             120             125
227
228        Glu Tyr Leu Cys Val Lys Ala Met Ile Leu Leu Asn Ser Ser Met Tyr
229             130             135             140
230
231        Pro Leu Val Thr Ala Thr Gln Asp Ala Asp Ser Ser Arg Lys Leu Ala
232        145             150             155             160
233
234        His Leu Leu Asn Ala Val Thr Asp Ala Leu Val Trp Val Ile Ala Lys
235                     165             170             175
236
237        Ser Gly Ile Ser Ser Gln Gln Gln Ser Met Arg Leu Ala Asn Leu Leu
238                     180             185             190
239
240        Met Leu Leu Ser His Val Arg His Ala Ser Asn Lys Gly Met Glu His
241             195             200             205
242
243        Leu Leu Asn Met Lys Cys Lys Asn Val Val Pro Val Tyr Asp Leu Leu
244             210             215             220
245
246        Leu Glu Met Leu Asn Ala His Val Leu
247        225             230
248
249  (2) INFORMATION FOR SEQ ID NO: 5:
250
251        (i) SEQUENCE CHARACTERISTICS:
252            (A) LENGTH: 477 amino acids
253            (B) TYPE: amino acid
254            (C) STRANDEDNESS: single
255            (D) TOPOLOGY: unknown
256
257     (ii) MOLECULE TYPE: protein
258
```

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION   *US/08/826,361A*

INPUT SET: S21390.raw

Line   Error   Original Text